THE GOOD, THE BAD, AND THE UGLY OF PREDICTIVE SCIENCE

François M. Hemez[1]	Yakov Ben-Haim[2]
Engineering Sciences and Applications	*Faculty of Mechanical Engineering*
Los Alamos National Laboratory	*Technion — Israel Institute of Technology*
Los Alamos, New Mexico 87545, U.S.A.	*Haifa 32000 Israel*

Extended Abstract

In computational physics and engineering, numerical models are developed to predict the behavior of a system whose response cannot be measured experimentally. A key aspect of science-based predictive modeling is the assessment of prediction credibility. Credibility, which is usually demonstrated through the activities of model Verification and Validation (V&V), quantifies the extent to which simulation results can be analyzed with confidence to represent the phenomenon of interest with an accuracy consistent with the intended use of the model [1].

The paper develops the idea that assessing the credibility of a mathematical or numerical model must combine three components: 1) Improving the fidelity to test data; 2) Studying the robustness of prediction-based decisions to variability, uncertainty, and lack-of-knowledge; and 3) Establishing the expected prediction accuracy of the models in situations where test measurements are not available. A Theorem demonstrates the irrevocable trade-off between *"the Good, the Bad, and the Ugly,"* or robustness-to-uncertainty, fidelity-to-data, and confidence-in-prediction.

1. Fidelity, Robustness, and Confidence

Even though the conventional activities of model V&V are generally restricted to improving the fidelity-to-data through the correlation of test and simulation results and the calibration of model parameters [2, 3], the other two components are equally important. The main reason is that optimal models—in the sense of models that minimize the prediction errors with respect to the available test data—possess exactly zero robustness to uncertainty and lack-of-knowledge [4]. This means that small variations in the setting of model parameters, or small errors in the knowledge of the functional form of the models, can lead to an actual fidelity that is significantly poorer than the one demonstrated through calibration.

Clearly, fidelity-to-data matters because no analyst will trust a numerical simulation that does not reproduce the measurements of past experiments or historical databases. Robustness-to-uncertainty is equally critical to minimize the vulnerability of decisions to uncertainty and lack-of-knowledge. It may be argued, however, that the most important aspect of credibility is the assessment of confidence-in-prediction, which is generally not addressed in the literature.

Assessing the confidence-in-prediction here refers to an assessment of prediction error away from settings where physical experiments have been performed, which must include a rigorous quantification of the sources of variability, uncertainty, and lack-of-knowledge, and their sensitivity effects on model prediction. The concepts of fidelity-to-data, robustness-to-

[1] Technical Staff Member, ESA-WR; Ph.: 505-663-5204; E-mail: hemez@lanl.gov.
[2] Yitzhak Moda'i Chair in Technology and Economics; Ph.: (+972) 4-829-3262; E-mail: yakov@technion.ac.il.

Approved for unlimited, public release on June 30, 2003 *LA-UR-03-4283*, ***Unclassified***

uncertainty, and prediction confidence are illustrated in Figure 1. The numerical model is represented conceptually as a "black-box" input-output relationship between the inputs $(p_1;p_2)$ and output prediction y:[3]

$$y = M(p_1;p_2) \tag{1}$$

A domain such as $[p_1^{(min)};p_1^{(max)}]$ x $[p_2^{(min)};p_2^{(max)}]$ represents the design space over which predictions must be obtained. Such requirement implies that the prediction accuracy must be established for all settings $(p_1;p_2)$ in the design domain $[p_1^{(min)};p_1^{(max)}]$ x $[p_2^{(min)};p_2^{(max)}]$.

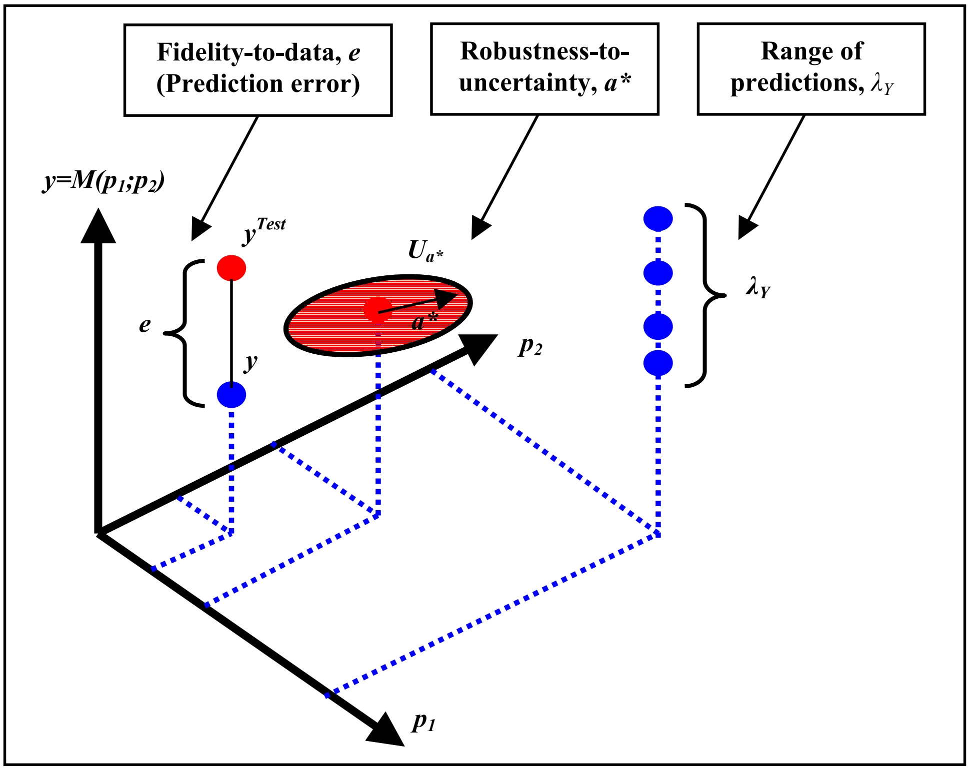

Figure 1. Concepts of fidelity-to-data, robustness, and prediction accuracy.

Fidelity-to-data represents the "distance" e—assessed with the appropriate metrics, possibly statistical tests if probabilistic information is involved—between physical measurements y^{Test} and simulation predictions y at a given setting $(p_1;p_2)$:

$$e = \left\| y^{Test} - y \right\| \tag{2}$$

Fidelity-to-data is pictured in Figure 1 as the vertical distance between a measurement y^{Test} and a prediction y for a physical experiment and a numerical simulation performed at the same setting $(p_1;p_2)$.

Robustness-to-uncertainty here refers to the range of settings $(p_1;p_2)$ that provide no more than a given level of prediction error e_{Max}. The concept of robustness-to-uncertainty a^* is formulated mathematically as:

[3] The input parameters $(p_1;p_2)$ represent settings such as, for example, the angle of attack and flow velocity of an aero-elastic simulation that predicts a coefficient of lift $y=C_L$. Another example would be the simulation of the response of a building to an Earthquake excitation, where the input parameters would represent the amplitude and frequency contents of the excitation and the output prediction would be structural stress levels.

$$\alpha^* = \max\left\{ \alpha > 0 \quad \text{such that} \quad \left\| y^{Test} - y \right\| \le e_{Max}, \quad \text{for all } (p_1; p_2) \in U_\alpha \right\} \tag{3}$$

where e_{Max} is a prediction error threshold not to be exceeded, U_a is a subset of the design domain $[p_1^{(min)};p_1^{(max)}] \times [p_2^{(min)};p_2^{(max)}]$ within which the parameters $(p_1;p_2)$ can vary, and a represents the size of the domain U_a. The family of domains U_a for $a>0$ can be arbitrarily defined, for example, as hyper-intervals centered about a nominal setting $(p_1^{(0)};p_2^{(0)})$. The only constraint to satisfy is that increasing values of the robustness parameter a must define nested domains U_a. It is emphasized that such definition can accommodate many models and representations of uncertainty and lack-of-knowledge.[4] The significance of the definition (3) is that all predictions made for settings $(p_1;p_2)$ included in U_{a^*} are guaranteed not to exceed the error threshold e_{Max}. The concept of robustness-to-uncertainty a^* is illustrated in Figure 1 by showing a domain U_a for which any prediction will not exceed e_{Max}.

Finally, the symbol λ_Y in Figure 1 refers to the range of predictions made by a family of potentially different models. The importance of λ_Y stems from the fact that, to have confidence in predictions, there should be as much consistency as possible between the predictions of equally credible "models" or sources of information.[5] Confidence in predictions is generally increased when different sources of evidence all reach the same conclusion. The concept of confidence-in-prediction is illustrated in Figure 1 by showing a range λ_Y of predictions obtained when different models are exercised to make predictions at a setting $(p_1;p_2)$ where no test data are available. In the remainder, the relationship between robustness-to-uncertainty a^*, fidelity-to-data e, and confidence-in-prediction or, equivalently, range-of-predictions λ_Y is discussed.

2. Total Uncertainty

Because the ultimate goal of model validation is to bound the predictive confidence by estimating a range of predictions λ_Y, the origin of evidence used in such assessment must be briefly discussed. In any realistic application, sources of evidence include expert judgment, back-of-the-envelope calculations, experimental measurements, or predictions obtained from phenomenological models or high-fidelity simulations. Together they define a knowledge space whose size is related to the consistency between the different sources of evidence. Figure 2 illustrates the concept of knowledge space where, for example, observations are obtained through two physical experiments (labeled *"Test 1"* and *"Test 2"*), opinions are collected from two experts (labeled *"Expert A"* and *"Expert B"*), and predictions are made by two models that incorporate different levels of fidelity to the physics.

The "size" of the knowledge domain illustrated in Figure 2 is an important component of predictive science because it defines the total uncertainty. Total uncertainty includes the effect of variability, uncertainty, and lack-of-knowledge of the physical phenomenon about which one is trying to make predictions. If all sources of evidence are consistent with each other, the total

[4] A first example is a probabilistic model of variability where the values of coefficients in the covariance matrix are controlled by the parameter a. A second example is a possibility structure π defined to represent a lack-of-knowledge, where the size of intervals is proportional to the parameter a. A third example is a family of fuzzy membership functions defined to represent expert judgment and linguistic ambiguity, where the membership functions are parameterized by a.

[5] As discussed in the next paragraph, "models" is here defined in a loose sense. Models can refer to physical models, numerical models, and other sources of information such as expert opinion and indirect measurements obtained when the data collected during physical experiments are interpreted through models.

uncertainty is expected to be small. Conversely, a large value of the total uncertainty metric tends to indicate a significant lack-of-knowledge. Although a formal relationship has not currently been developed, there is an obvious connection between the concept of total uncertainty and the notion of confidence-in-prediction or, equivalently, with the range-of-predictions λ_Y. Metrics are currently being studied to appropriately measure the total uncertainty represented by different sources of evidence [5]. Even though it is usually restricted to parametric variability, model output sensitivity should also be concerned with studying the effects of epistemic uncertainty and lack-of-knowledge.

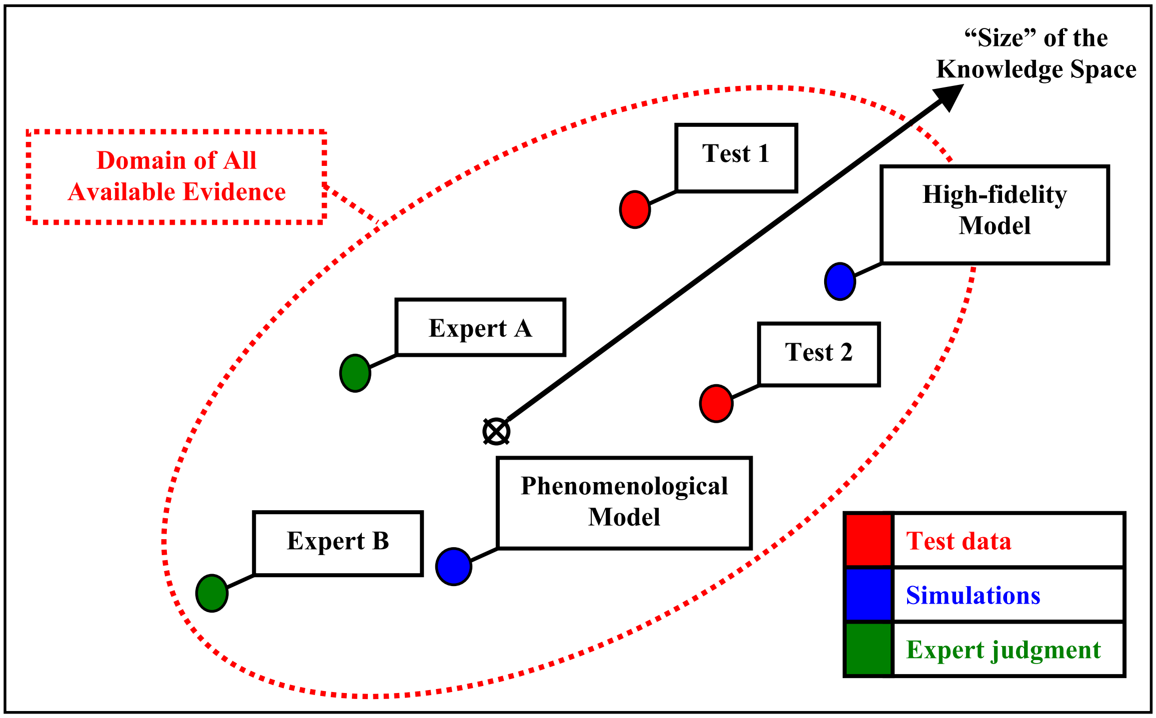

Figure 2. Concept of knowledge space.

3. Theoretical Results

The paper explores the trade-offs between fidelity-to-data e, robustness-to-uncertainty α, and range-of-predictions λ_Y. We prove a Theorem that can be summarized by:

$$\frac{\partial \lambda_Y}{\partial \alpha} \geq 0 \tag{4}$$

which means that a revision of the model, with the purpose of enhancing robustness to modeling error, also increases the range of predictions. In other words, robustness and prediction confidence are antagonistic attributes of any model. The proof relies on the information-gap description of uncertainty and no restrictive assumption is made regarding the models, sources or types of uncertainty, and their mathematical representations [6].

The Theorem can be further extended to prove the following three inequalities:

$$\frac{\partial \alpha}{\partial e} \geq 0, \quad \frac{\partial \lambda_Y}{\partial \alpha} \geq 0, \quad \frac{\partial \lambda_Y}{\partial e} \geq 0 \tag{5}$$

which express the three trade-offs between fidelity, robustness, and confidence-in-prediction:

- *Robustness decreases as fidelity improves.* The robustness α gets larger if the prediction error e gets larger. Numerical simulations made to better reproduce the test data become more vulnerable to errors in modeling assumptions, errors in the functional form of the model, and uncertainty and variability in the model parameters.
- *The range of predictions increases as robustness improves.* The range of predictions λ_Y gets larger if the robustness α gets larger. Numerical simulations that are more immune to uncertainty and modeling errors provide a wider, hence less consistent, range of predictions.
- *The range of predictions increases as fidelity improves.* The range of predictions λ_Y gets larger if the prediction error e gets larger. Numerical simulations made to better reproduce the available test data provide a smaller range of predictions, hence, enhancing the consistency between the predicted responses. Although intuitive, this result is not necessarily a good thing when the models are employed to analyze configurations of the system that are very different from those tested.

These trade-offs imply that it is not possible to have, simultaneously, high fidelity, large robustness, and large confidence-in-prediction. High fidelity (small e) implies that the model is true to the measurements, which adds warrant to the model. Large robustness (large α) strengthens belief in the validity of the model or family of models. Consistent predictions (small λ_Y) imply that the models that are equivalent in terms of fidelity, also agree in their predictions of the system behavior.

The conflict between robustness, fidelity, and confidence-in-prediction is reminiscent of Hume's critique of empirical induction. Our analysis shows that past measurements, accompanied by incomplete understanding of the measured process, cannot unequivocally establish true predictions of the behavior of the system.

4. Application

The theoretical results are illustrated with an engineering application whose purpose is to predict the peak acceleration transmitted through a layer of non-linear hyper-foam material subjected to a transient impact [7]. Evidence is obtained by analyzing low-fidelity models, analyzing high-fidelity finite element models, eliciting expert opinion, and collecting physical measurements. These sources of information are illustrated in Figure 3 that shows the ranges of peak acceleration values expected to result from the propagation of the shock wave.

The decision that must be supported by this analysis is to assess whether or not the peak acceleration value will exceed a critical level. Uncertainty about the information shown in Figure 3 is represented using a possibility distribution because it is epistemic in nature. A non-probabilistic sensitivity analysis is performed next to assess how the lack-of-knowledge about the physics of the shock transmission affects the safety margin, defined as the difference between the peak acceleration and a level not to be exceeded. The sensitivity is defined as the slope of the robustness-to-uncertainty, and it expresses the vulnerability of the safety margin to total uncertainty. Results are compared to a probabilistic-based reliability analysis to illustrate the consequence of not accounting for the lack-of-knowledge.

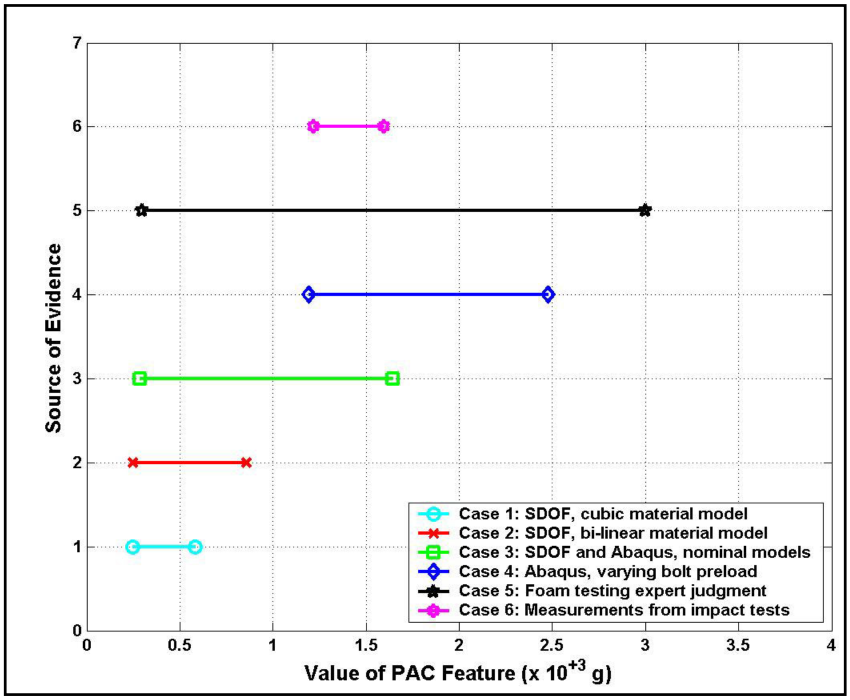

Figure 3. Study of the propagation of a shock wave through a hyper-foam material.

References

[1] Doebling, S.W., "Structural Dynamics Model Validation: Pushing the Envelope," *International Conference on Structural Dynamics Modeling: Test, Analysis, Correlation and Validation*, Madeira Island, Portugal, June 3-5, 2002.

[2] Mottershead, J.E., Friswell, M.I., "Model Updating in Structural Dynamics: A Survey," *Journal of Sound and Vibration*, Vol. 162, No. 2, 1993, pp. 347-375.

[3] Hemez, F.M., Doebling, S.W., "Inversion of Structural Dynamics Simulations: State-of-the-art and Orientations of the Research," *25th ISMA International Conference on Noise and Vibration Engineering*, Leuven, Belgium, Sept. 13-15, 2000, pp. 403-413.

[4] Ben-Haim, Y., "Robust Rationality and Decisions Under Severe Uncertainty," *Journal of the Franklin Institute*, Vol. 337, 2000, pp. 171-199.

[5] Anderson, M.C., Booker, J., Hemez, F.M., Joslyn, C., Reardon, B., Ross, T., "Quantifying Total Uncertainty in a Validation Assessment Using Different Mathematical Theories," *14th Biennial Nuclear Explosives Design Physics Conference*, Los Alamos National Laboratory, Los Alamos, New Mexico, October 20-24, 2003.

[6] Ben-Haim, Y., **Information-Gap Decision Theory: Decisions Under Severe Uncertainty**, Academic Press, 2001.

[7] Hemez, F.M., Wilson, A.C., Doebling, S.W., "Design of Computer Experiments for Improving an Impact Test Simulation," *19th International Modal Analysis Conference,* Kissimmee, FL, Feb. 5-8, 2001, pp. 977-985.